

Problem Set 2 Knitted

Jiaxuan Liu

2024-10-18

Problem Set 2 Report

Introduction

1. “This submission is my work alone and complies with the 30538 integrity policy.” Add your initials to indicate your agreement: **JL**
2. “I have uploaded the names of anyone I worked with on the problem set here” asdgla;sdgnawe;oe
3. Late coins used this pset: ****__**** Late coins left after submission: ****__****

Q1: Counting NA values

For each column, how many rows are NA? Write a function that returns a two-column DataFrame where each row is a variable, the first column of the DataFrame is the name of each variable, and the second column of the DataFrame is the number of times that the column is NA.

Importing the dataset

```
import matplotlib.pyplot as plt
import jupyter
import pandas as pd

# File path for the parking tickets dataset
file_path = '/Users/doudouliu/Desktop/python 2/student30538/problem_sets/ps2/data/parking_tickets.csv'

# Reading the CSV file
```

```
df = pd.read_csv(file_path)
```

```
# Display first few rows to ensure the dataset is loaded correctly
print(df.head())
```

	Unnamed: 0	ticket_number	issue_date	violation_location	\
0	1	51482901.0	2007-01-01 01:25:00	5762 N AVONDALE	
1	2	50681501.0	2007-01-01 01:51:00	2724 W FARRAGUT	
2	3	51579701.0	2007-01-01 02:22:00	1748 W ESTES	
3	4	51262201.0	2007-01-01 02:35:00	4756 N SHERIDAN	
4	5	51898001.0	2007-01-01 03:50:00	7134 S CAMPBELL	

	license_plate_number	license_plate_state	\
0	d41ee9a4cb0676e641399ad14aaa20d06f2c6896de6366...	IL	
1	3395fd3f71f18f9ea4f0a8e1f13bf0aa15052fc8e5605a...	IL	
2	302cb9c55f63ff828d7315c5589d97f1f8144904d66eb3...	IL	
3	94d018f52c7990cea326d1810a3278e2c6b1e8b44f3c52...	IL	
4	876dd3a95179f4f1d720613f6e32a5a7b86b0e6f988bf4...	IL	

	license_plate_type	zipcode	violation_code	\
0	PAS	606184118.0	0964090E	
1	PAS	606454911.0	0964090E	
2	PAS	604116803.0	0964150B	
3	PAS	606601345.0	0976160F	
4	PAS	606291432.0	0964100A	

	violation_description	...	fine_level2_amount	\
0	RESIDENTIAL PERMIT PARKING	...	100	
1	RESIDENTIAL PERMIT PARKING	...	100	
2	PARKING/STANDING PROHIBITED ANYTIME	...	100	
3	EXPIRED PLATES OR TEMPORARY REGISTRATION	...	100	
4	WITHIN 15' OF FIRE HYDRANT	...	200	

	current_amount_due	total_payments	ticket_queue	ticket_queue_date	\
0	0.0	50.0	Paid	2007-03-20	
1	0.0	50.0	Paid	2007-01-31	
2	122.0	0.0	Notice	2007-02-28	
3	0.0	50.0	Paid	2007-01-11	
4	0.0	100.0	Paid	2007-04-25	

	notice_level	hearing_disposition	notice_number	officer	\
0	DETR	Liabile	5.080059e+09	17266	

1	VIOL	NaN	5.079876e+09	10799
2	SEIZ	NaN	5.037862e+09	17253
3	NaN	NaN	5.075310e+09	3307
4	DETR	NaN	5.073568e+09	16820

	address
0	5700 n avondale, chicago, il
1	2700 w farragut, chicago, il
2	1700 w estes, chicago, il
3	4700 n sheridan, chicago, il
4	7100 s campbell, chicago, il

[5 rows x 24 columns]

```
/var/folders/b0/6kybjmv57sv8drq7v4kr7w240000gn/T/ipykernel_84587/3451754831.py:9: DtypeWarning
df = pd.read_csv(file_path)
```

#Q1a, How many NAs

```
# Function to count NAs for each column
Nna = df.isna().sum()
print(Nna)
```

Unnamed: 0	0
ticket_number	0
issue_date	0
violation_location	0
license_plate_number	0
license_plate_state	97
license_plate_type	2054
zipcode	54115
violation_code	0
violation_description	0
unit	29
unit_description	0
vehicle_make	0
fine_level1_amount	0
fine_level2_amount	0
current_amount_due	0
total_payments	0
ticket_queue	0

ticket_queue_date	0
notice_level	84068
hearing_disposition	259899
notice_number	0
officer	0
address	0
dtype: int64	